Noise Reduction of Undersampled ISAR Images by Entropy-Based Regularization

by
Brett Borden
Research and Technology Division

OCTOBER 1996

NAVAL AIR WARFARE CENTER WEAPONS DIVISION
CHINA LAKE, CA 93555-6100

Naval Air Warfare Center Weapons Division

FOREWORD

The research described in this report was performed during fiscal year 1996 as part of an effort to improve radar classification and identification capabilities for noncooperative airborne targets. This problem continues to be a primary goal of radar research programs and considerable effort has been expended within the last few decades in attempts to solve it. The accuracy of these methods depends on available radar resolution and signal-to-noise ratio.

A generalization to traditional maximum entropy techniques has been developed, which explicitly accounts for the possibility of noisy and complex-valued data. The current work has concentrated on demonstrating the applicability of the basic concepts and not upon developing computer-efficient implementations of the algorithm.

This effort was supported by Office of Naval Research, Code 313.

This report was reviewed for technical accuracy by Carey Schwartz.

Approved by
R. L. Derr, *Head*
Research and Technology Division
1 October 1996

Under authority of
J. V. CHENEVEY
RADM, U.S. Navy
Commander

Released for publication by
S. HAALAND
Director for Research and Engineering

NAWCWPNS Technical Publication 8319

Published by . . . . . Technical Information Division
Collation . . . . . Cover, 9 leaves
First printing . . . . . 25 copies

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 1996	Interim report—October 1995-September 1996

4. TITLE AND SUBTITLE
Noise Reduction of Undersampled ISAR Images by Entropy-Based Regularization

5. FUNDING NUMBERS
N00014-94-WX-30309

6. AUTHOR(S)
Brett Borden

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Air Warfare Center Weapons Division
China Lake, CA 93555-6100

8. PERFORMING ORGANIZATION REPORT NUMBER
NAWCWPNS TP 8319

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
W. J. Miceli
Office of Naval Research
Code 313
800 North Quincy Street
Arlington, VA 22217-5660

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12A. DISTRIBUTION/AVAILABILITY STATEMENT
A Statement; public release; distribution unlimited.

12B. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

(U) The research described in this report was performed during the 1996 fiscal year as part of an effort to improve radar classification and identification capabilities for noncooperative airborne targets. Effective radar-based imaging techniques should be able to employ the information contained in both the amplitude and the phase of the scattered field data. Moreover, since practical radar systems collect restricted and noisy data, image reconstruction algorithms should allow for the inclusion of *a priori* information about the target and should attempt to mitigate the effects of noise. All three issues—complex-valued data, prior information, and noise mitigation—are often not adequately treated by conventional entropy-based methods (and are mutually exclusive in many other approaches). We present a Bayesian regularization method that employs a cross-entropy functional and does not require the measured data to positive and real-valued. In fact, this technique is not restricted to a particular type of data and we present reconstruction examples for several different imaging problems. The basic model of this approach is similar to that used in usual maximum *a posteriori* analysis and allows for a more general relationship between the image and its "configuration entropy" than that usually employed.

14. SUBJECT TERMS		15. NUMBER OF PAGES
Radar Imaging Entropy Cross-Entropy	Regularization	16
		16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

INTRODUCTION

Recent advances in radar imaging systems have been cause for a renewed interest in appropriate image reconstruction algorithms. This problem is somewhat different from the more usual (optical) imaging situation because the radar data to be analyzed often consist of the in-phase and quadrature components of the scattered electromagnetic field, and so these data are *complex-valued.* Moreover, it is clear from the coherent nature of radar that the full amplitude and phase properties of the reconstruction problem must be addressed if the details of the radar/target interaction are to be deciphered.

Radar imaging shares with optical image reconstruction the two central problems of resolution degradation and sensitivity to noise (see, for example, References 1 through 3 and references cited therein). The resolution problem occurs when the data are limited, and overcoming this problem requires additional "data" be added in a way that either fills-in the unmeasured information or, at least, precludes unwanted artifacts. Maximum entropy (or minimum cross-entropy) methods deal neatly with these information-oriented issues and enjoy considerable popularity (References 4 and 5). Unfortunately, while there has been some effort to deal with noise or complex-valued data (Reference 6), traditional entropy-based approaches are often inadequate.

Noise moderation is generally accomplished by more general Bayesian techniques. Here, the noise effects and the additional (unmeasured) data are separately modeled and accounted for. In addition, because of the generality of Bayesian methods, complex-valued data presents no fundamental complications and are often employed.

We will examine the issues of noise, information entropy, and complex-valued data within the context of Bayesian image reconstruction. We will do this by considering the problem of regularization of ill-posed linear operator problems using Bayes risk functionals. After reviewing the basic concepts and establishing notation we will introduce a Markov random field (MRF) image model that will allow us to account for the possibility of extended image elements. Our basic result is a regularization functional that generalizes traditional maximum entropy approaches to the case of noisy complex-valued data. A simple image reconstruction algorithm, which applies these results to the MRF model, is presented as well as an example relevant to inverse synthetic aperture array (ISAR) imaging.

ILL-POSED PROBLEMS

For many systems, a measurement obtained at position $\mathbf{x}_i$ can be modeled as a linear operation by a kernel K on an object function f:

$$F(\mathbf{x}_i) = \int_D K(\mathbf{x}_i, \mathbf{y}) f(\mathbf{y})\, d\mathbf{y}. \tag{1}$$

The goal in object reconstruction is to estimate $f(\mathbf{y})$ from measured data $\{F(\mathbf{x}_i)\}_{i=1,\ldots,N_1 N_2}$ when K is known. This is an inverse problem and considerable effort has been expended in developing reliable and efficient algorithms for its solution in a wide variety of applications.

When the measurements $\{F(\mathbf{x}_i)\}$ can be made with perfect accuracy, then the inversion problem is complicated only by the fact that the measurements are necessarily limited in extent and/or finite in number. If f is considered to belong to a Hilbert space, then Equation 1 can be written as $\mathbf{F} = \mathbf{K}\mathbf{f}$, where $\mathbf{K}$ is the linear operator corresponding to K that maps $\mathbf{f}$ to the measurements $\mathbf{F}$.

Unfortunately, most measurements will be contaminated by noise so that Equation 1 becomes

$$\mathbf{F} = \mathbf{K}\mathbf{f} + \mathbf{n} \tag{2}$$

where $\mathbf{n}$ represents a random noise component. When the problem is ill-posed, the inverse of $\mathbf{K}$ may map small variations in the data to large (and unwarranted) variations in $\mathbf{f}$. In general, ill-posedness will occur whenever $\mathbf{K}$ is compact—and in particular, whenever K is square-integrable and nondegenerate. Consequently, the inverse problem associated with Equation 1 is *often* ill-posed and the effects of this noise contamination can be significant.

A traditional approach for obtaining an estimate $\hat{\mathbf{f}}$ of f is to choose the $\mathbf{g}$, from the set Φ of all possible solutions, that minimizes the noise error "energy":

$$\|\mathbf{F} - \mathbf{K}\hat{\mathbf{f}}\|^2 = \min_{\mathbf{g}\in\Phi} \|\mathbf{F} - \mathbf{K}\mathbf{g}\|^2. \tag{3}$$

This method makes sense even when $\mathbf{K}$ does not have a proper inverse (for example, when $\mathbf{K}$ is not square). Moreover, the solution process implied by Equation 3 can frequently be implemented in a straightforward and computationally efficient manner. The generalized inverse implied by this "least-squares" solution, however, is unbounded (unless $\mathbf{K}$ has finite rank) and so the effects of the noise component in the data are generally unmoderated by ordinary least-squares analysis. Successfully addressing the issue of noise in ill-posed inverse problems requires a detailed understanding of both the noise process and the set Φ of acceptable solutions. It is known, for example, that by sufficiently restricting Φ, we may be able to "regularize" the inversion process and control the effects of noise.

REGULARIZATION

Because $\mathbf{n}$ is random, the measured data will be a realization selected from some infinitely large ensemble and we can only obtain the estimate $\hat{\mathbf{f}}$ in a statistical sense. Consequently, our knowledge of $\mathbf{f}$ will come from a probability density $p(\mathbf{f})$. In fact, the density $p(\mathbf{f})$ *is* the estimated "solution" to the problem because it describes what can be known about $\mathbf{f}$, and once $p(\mathbf{f})$ is determined an estimate $\hat{\mathbf{f}}$ can be obtained in a variety of ways.

Moderating the effects of noise is not the only problem associated with object reconstruction: often we must also deal with the limitations imposed by incomplete measurements. To see this, consider Equation 2 and let $\mathbf{k}_i$ denote the i^{th} row of $\mathbf{K}$. Since the data obtained by Equation 1 are necessarily discrete and finite, only those components of $\mathbf{f}$ that lie in the subspace spanned by the set of all $\mathbf{k}_i$ contribute to the measurements. This subspace is the *measurement space* and components of $\mathbf{f}$ that lie in the subspace orthogonal to the measurement space (the *null space*) will not affect the measurement. Estimates of $\mathbf{f}$ of the form of Equation 3 may display unwanted artifacts because they make no effort to account for the null space components of $\mathbf{f}$.

A statistical estimation method, which generalizes the idea behind Equation 3, sets the estimate $\hat{\mathbf{f}}$ of $\mathbf{f}$ to minimize the *Bayes risk*:

$$B(\hat{\mathbf{f}}) = \int C(\mathbf{f}, \hat{\mathbf{f}})\; p_{\mathbf{f}|\mathbf{F}}(\mathbf{f}|\mathbf{F})\; d\mathbf{f}\,, \tag{4}$$

where $C(\mathbf{f}, \hat{\mathbf{f}})$ is the cost function representing some difference error between $\mathbf{f}$ and $\hat{\mathbf{f}}$, and $d\mathbf{f}$ denotes an appropriate measure on the space of all $\mathbf{f}$. The function $p_{\mathbf{f}|\mathbf{F}}(\mathbf{f}|\mathbf{F})$ is the posterior distribution of $\mathbf{f}$ given the known values $\mathbf{F}$:

$$p_{\mathbf{f}|\mathbf{F}}(\mathbf{f}|\mathbf{F}) \equiv \frac{p_{\mathbf{F}|\mathbf{f}}(\mathbf{F}|\mathbf{f}) p_{\mathbf{f}}(\mathbf{f})}{p_{\mathbf{F}}(\mathbf{F})}. \tag{5}$$

The Bayes risk is the expected cost of an error in estimation and Bayesian estimation seeks to minimize this expected cost.

Several cost functions have become popular. $C_{\mathrm{V}}(\mathbf{f}, \hat{\mathbf{f}}) = \|\mathbf{f} - \hat{\mathbf{f}}\|^2$ is the squared error and yields the minimum variance estimate:

$$\begin{aligned} \frac{\partial B_{\mathrm{V}}(\hat{\mathbf{f}})}{\partial \hat{\mathbf{f}}} = \frac{\partial}{\partial \hat{\mathbf{f}}} \int \|\mathbf{f} - \hat{\mathbf{f}}\|^2\; p_{\mathbf{f}|\mathbf{F}}(\mathbf{f}|\mathbf{F})\; d\mathbf{f} = 2 \int (\mathbf{f} - \hat{\mathbf{f}})\; p_{\mathbf{f}|\mathbf{F}}(\mathbf{f}|\mathbf{F})\; d\mathbf{f} = 0 \\ \Rightarrow \quad \hat{\mathbf{f}}_{\mathrm{MV}} = \int \mathbf{f}\; p_{\mathbf{f}|\mathbf{F}}(\mathbf{f}|\mathbf{F})\; d\mathbf{f}. \end{aligned} \tag{6}$$

Another frequently employed cost function sets

$$C_{\mathrm{UCF}}(\mathbf{f},\hat{\mathbf{f}}) = \begin{cases} 0 & \text{if } \|\mathbf{f}-\hat{\mathbf{f}}\| < \frac{\epsilon}{2} \\ \frac{1}{\epsilon} & \text{if } \|\mathbf{f}-\hat{\mathbf{f}}\| \geq \frac{\epsilon}{2}. \end{cases} \tag{7}$$

This is the uniform cost function and, in the limit $\epsilon \rightarrow 0$, leads to the maximum *a posteriori* estimate:

$$\lim_{\epsilon \rightarrow 0} \frac{\partial B_{\mathrm{UCF}}(\hat{\mathbf{f}})}{\partial \hat{\mathbf{f}}} = 0 \quad \Rightarrow \quad \hat{\mathbf{f}}_{\mathrm{MAP}} = \arg\max\, p_{\mathbf{f}|\mathbf{F}}(\mathbf{f}|\mathbf{F}). \tag{8}$$

When the cost function is symmetric and convex about its minimum, and $p_{\mathbf{f}|\mathbf{F}}(\mathbf{f}|\mathbf{F})$ is symmetric about $\hat{\mathbf{f}}_{\mathrm{MV}}$, then it can be shown that the optimal Bayes estimate is equal to the minimum variance estimate (Reference 7). Consequently, the minimum variance estimate is *the* optimal estimator for a wide variety of very reasonable circumstances.

Of course, this kind of estimation methodology presumes that the specification of $p_{\mathbf{f}|\mathbf{F}}(\mathbf{f}|\mathbf{F})$ in Equation 5 is sufficiently nontrivial that the application of Equation 6 or 8 is relevant to the problem of moderating the effects of limited and corrupted data. This can only be accomplished when the prior density $p_{\mathbf{f}}(\mathbf{f})$ includes genuine information about $\mathbf{f}$ that is not already contained in the measured data. Moreover, the estimates of Equations 6 and 8 reflect only limited properties of the "complete solution" $p_{\mathbf{f}|\mathbf{F}}(\mathbf{f}|\mathbf{F})$: $\hat{\mathbf{f}}_{\mathrm{MV}}$ is the first moment of the density $p_{\mathbf{f}|\mathbf{F}}$ while $\hat{\mathbf{f}}_{\mathrm{MAP}}$ is its maximum. A real practical difficulty in obtaining any estimate $\hat{\mathbf{f}}$ lies in first obtaining a meaningful estimate of $p_{\mathbf{f}|\mathbf{F}}(\mathbf{f}|\mathbf{F})$.

The determination of $p_{\mathbf{f}|\mathbf{F}}(\mathbf{f}|\mathbf{F})$ requires the specification of the factors $p_{\mathbf{F}|\mathbf{f}}(\mathbf{F}|\mathbf{f})$ and $p_{\mathbf{f}}(\mathbf{f})$. (The denominator $p_{\mathbf{F}}(\mathbf{F})$ can be considered to be a normalizing constant.) The first of these factors is the *a priori* probability density of the measurement given the object $\mathbf{f}$ and is usually assumed to be determined by the noise statistics $p_{\mathbf{n}}(\mathbf{n})$ of the measurement process. Because of Equation 2, we can set

$$p_{\mathbf{F}|\mathbf{f}}(\mathbf{F}|\mathbf{f}) = p_{\mathbf{n}}(\mathbf{F}-\mathbf{K}\mathbf{f}) \tag{9}$$

when $p_{\mathbf{n}}(\mathbf{n})$ is known. The second term, $p_{\mathbf{f}}(\mathbf{f})$, is the *a priori* probability density of $\mathbf{f}$. This is the term that must be specified in a non-trivial manner if the artifacts caused by insufficient data are to be ameliorated. Unfortunately, specifying $p_{\mathbf{f}}(\mathbf{f})$ in such a non-trivial way requires that considerable information about $\mathbf{f}$ be known in advance of the solution process and this is often not possible. A more common approach models $p_{\mathbf{f}}(\mathbf{f})$ to include only those properties known to be true for a large and general class of $\mathbf{f}$. This generality, however, often implies considerable ambiguity in actually specifying *a* $p_{\mathbf{f}}(\mathbf{f})$ to be used in a particular situation. Determining the "best" $p_{\mathbf{f}}(\mathbf{f})$ for the estimation problem at hand is often a significant additional problem in itself.

One approach to estimation of $p(\mathbf{f})$ begins by extending the allowed class of cost functions $C \rightarrow \tilde{C}$ to include differences in the "estimation error" of density functions. A common example of this is the cost function

$$\tilde{C}(p(\mathbf{f}), \hat{p}(\mathbf{f})) = \ln p(\mathbf{f}) - \ln \hat{p}(\mathbf{f}). \tag{10}$$

(The quantity $-\ln p(\mathbf{f})$ is known as the "self information" of $p(\mathbf{f})$ and weights the information "importance" of an event by its frequency—the less frequent event being the more important.)

When p and $\hat{p}$ both have exponential arguments in $\mathbf{f}$, then Equation 10 can be easily compared to Equation 4. The definition of Equation 10 is more general, however, and the negative of

$$\tilde{B}(p, \hat{p}) = \int \ln \left(\frac{p(\mathbf{f})}{\hat{p}(\mathbf{f})} \right) \, p(\mathbf{f}) \, d\mathbf{f} \tag{11}$$

is the so-called *cross-entropy*, or *Kullback "distance"* between p and $\hat{p}$. $\tilde{B}(p, \hat{p})$ measures the (non-symmetric) information difference between the two distributions and provides a mechanism for estimating our model error: of all the possible $\hat{p}$ that are consistent with the prior information, choose the one that minimizes the information difference to p.

As with noise-corrupted measurements, the deleterious effects caused by restricted measurements can be moderated by including an accurate prior model. When such additional information about $\mathbf{f}$ is known in advance of the measurements (for example, that an acceptable f must be "smooth" or that $\mathbf{f}$ has a known support) then the quality of the estimated solution may often be dramatically improved by including it. This is the idea behind regularization: rather than force a solution to the original ill-posed problem, select instead the solution to a corresponding well-posed problem that is "close" to the original problem. The framework for this approach was established in Reference 2 and usually proceeds by seeking estimates that are equally close to conflicting criteria. For example, if $\mathbf{f}_0$ is a *prior* estimate of object parameters incorporating "general" known properties of $\mathbf{f}$ (including its null space components), then our estimate should weigh and balance this information against the estimate that best fits the measured data.

In terms of our original problem of Equation 2, let $B_1(\mathbf{f}, \mathbf{F})$ denote a measure of the "difference" between $\mathbf{f}$ and the solution implied by the measured data. Similarly, let $B_2(\mathbf{f}, \mathbf{f}_0)$ measure the difference between $\mathbf{f}$ and the *a priori* information (expressed symbolically by the parameter set $\mathbf{f}_0$). Regularization methods seek an estimate $\hat{\mathbf{f}}$ that minimizes the functional

$$B(\mathbf{f}) = B_1(\mathbf{f}; \mathbf{F}) + \lambda B_2(\mathbf{f}; \mathbf{f}_0), \qquad \mathbf{f} \in \Phi. \tag{12}$$

(Here $\lambda > 0$ is the regularization parameter and controls the relative influence of each term.)

A recognized difficulty with applying Equation 12 is in the choice of appropriate and related difference functionals. This is because $B_1(\mathbf{f}, \mathbf{F})$ relates an estimate $\hat{\mathbf{f}}$ to the data $\mathbf{F}$ whereas, based

on our discussion, $B_2(\mathbf{f}, \mathbf{f}_0)$ must often relate an estimate of $p_{\mathbf{f}|\mathbf{F}}(\mathbf{f}|\mathbf{F})$ to $\hat{p}(\mathbf{f})$—and these are *fundamentally* different quantities. One approach to this problem is to model $p(\mathbf{f}_i) \propto \mathbf{f}_i$ and use the atomic measure over D so that integrals become sums. This is the so-called "configuration density." Applied to Equation 11, for example, this choice of density results in the configuration entropy difference.

The configuration entropy has been successfully used in a variety of problems and its appropriateness to these problems has been examined and debated. Independent of these concerns, however, is the question of how to generalize the method to include an $\mathbf{f}$ with non-positive, or even complex-valued, components. Various schemes have tried $p(\mathbf{f}_i) \propto |\mathbf{f}_i|$ and $p(\mathbf{f}_i) \propto \text{Re}\, \mathbf{f}_i$, but these are generally *ad hoc* models and do not enjoy a universal appeal. A better approach, perhaps, is to use a $B_1(\mathbf{f}, \mathbf{F})$ which is more compliant with the variation of $p_{\mathbf{f}|\mathbf{F}}(\mathbf{f}|\mathbf{F})$.

One way of doing this is to link a specific estimate to the data up-front. For example,

$$B_1(p_\mathbf{f}) = \int \|\mathbf{F} - \mathbf{K}\mathbf{f}\|^2 \, p_{\mathbf{f}|\mathbf{F}}(\mathbf{f}|\mathbf{F}) \, d\mathbf{f}. \tag{13}$$

If we use the cross-entropy functional (Equation 11) for $B_2(\mathbf{f}, \mathbf{f}_0)$, then Equation 12 becomes

$$B(p_\mathbf{f}) = \int \|\mathbf{F} - \mathbf{K}\mathbf{f}\|^2 \, p_{\mathbf{f}|\mathbf{F}}(\mathbf{f}|\mathbf{F}) \, d\mathbf{f} + \lambda \int \ln\left(\frac{p_{\mathbf{f}|\mathbf{F}}(\mathbf{f}|\mathbf{F})}{p_\mathbf{f}(\mathbf{f};\mathbf{f}_0)}\right) \, p_{\mathbf{f}|\mathbf{F}}(\mathbf{f}|\mathbf{F}) \, d\mathbf{f} \tag{14}$$

It is known that a selection rule that is both "regular" and "local" must be of the form of a functional that minimizes the distance between the estimated solution and some prior guess (Reference 8). Using "composition consistency" arguments, it is shown in Reference 8 that selection rules for general $\mathbf{f}$ (i.e., not having strictly positive components) *must* be of the least-squares type. It is also shown (References 8–10) that composition consistent selection rules for $p(\mathbf{f})$ (with strictly positive components) *must* be of the cross-entropy type. Consequently, our choice of regularization functional (Equation 14) is regular, local, and composition consistent. Moreover, given that our "distances" are formed by comparing $\mathbf{K}\mathbf{f}$ to $\mathbf{F}$ and by comparing $p_{\mathbf{F}|\mathbf{f}}$ to $p_\mathbf{f}$ (and while enforcing a connection between $p_{\mathbf{F}|\mathbf{f}}$ and $\hat{\mathbf{f}}$), we believe that the choice of Equation 14 is a natural one, and we shall use it for the remainder of this discussion.

OBJECT FUNCTION MODELS

Because of Equation 2, the statistics of $\mathbf{n}$ will determine $p_{\mathbf{F}|\mathbf{f}}$. However, the choice for $p_\mathbf{f}$ should be based upon a general understanding of the possible behavior of the object independent of the

measurement process (References 11–17). If $p_{\mathbf{f}}$ is modeled as an M-dimensional Gaussian with $\mathbf{f}_0 = \{\mathbf{R_f}, \bar{\mathbf{f}}\}$ so that

$$p_{\mathbf{f}}(\mathbf{f};\mathbf{f}_0) = \frac{1}{(2\pi)^{M/2}\sqrt{\det \mathbf{R_f}}} \exp(-\frac{1}{2}(\mathbf{f}-\bar{\mathbf{f}})^{\dagger}\mathbf{R_f}^{-1}(\mathbf{f}-\bar{\mathbf{f}})), \tag{15}$$

($(^{\dagger})$ denotes complex-conjugate transpose), then it is essential that $\mathbf{R_f}^{-1}$ be chosen to represent the known structure in the object function: the choice $\mathbf{R_f}^{-1} = \mathbf{I}/\sigma^2$ is equivalent to saying that *nothing* is known about the structural possibilities of $\mathbf{f}$, except the initial guess $\bar{\mathbf{f}}$.

In order to model extended object structure (for example, discontinuities) while still accounting for noise, we need to be able to distinguish between random variations attributable to signal contamination and "jumps" associated with the behavior of f over a domain or "neighborhood." Specifically, a neighborhood $\partial\mathbf{u} \in D$ of $\mathbf{u}$ is a set of points with the property that $\forall\, \mathbf{u}, \mathbf{v} \in D$, $\mathbf{u} \notin \partial\mathbf{u}$ and $\mathbf{v} \in \partial\mathbf{u} \iff \mathbf{u} \in \partial\mathbf{v}$. If $p(\mathbf{f})$ has the characteristic that

$$\forall\, \mathbf{u} \in D \quad p(f(\mathbf{u})|f(\mathbf{v}), \mathbf{v} \neq \mathbf{u}) = p(f(\mathbf{u})|f(\mathbf{v}), \mathbf{v} \in \partial\mathbf{u}), \tag{16}$$

then f is said to be an MRF. MRFs are useful as object models because they restrict computation to be local and are general enough to model the behavior of $f(\mathbf{u})$ in the neighborhood of $\mathbf{u}$. The MRF property applied to Equation 15, for example, implies that $\mathbf{R_f}^{-1}$ is a sparse matrix with zero values at locations that do not correspond to neighboring object elements.

In the case of Gaussian noise the Gaussian object model is often chosen because of its analytical advantage and the Gaussian Markov random field (GMRF) seems to offer the potential for including non-trivial extended prior object structure. When the object is *a priori* known to have discontinuities, however, the Gaussian model may be inappropriate because the squared difference $\|\mathbf{f}-\bar{\mathbf{f}}\|^2$ applies too high a penalty to these discontinuities (which will therefore be excessively smoothed). This drawback is not endemic to MRFs in general and there are many candidate prior object models that overcome the GMRF tendency to "over smooth," but that retain the more attractive features induced by concentrating on local object neighborhoods.

The GMRF is only one example of a much larger class of MRFs known as Gibbs distributions, which are defined over neighborhoods in D. A local set of points $c \subset D$ is known as a *clique* if $\forall\, \mathbf{u}, \mathbf{v} \in c$, $\mathbf{u}$ and $\mathbf{v}$ are neighbors. In general, a Gibbs distribution can be expressed in the form

$$p_g(\mathbf{f}) = \frac{1}{Z} \exp\left(-\int_{c\in C} V_c(\mathbf{f})\, dc\right), \tag{17}$$

where V_c is a "clique potential" function on D with the property that $V_c(\mathbf{f})$ depends only on the coordinates $\mathbf{u}$ of f which lie in c. The integral in Equation 17 is over the set C of all cliques in D, and $Z \equiv \int p_g(\mathbf{f})\, d\mathbf{f}$ is a normalization factor known as the "partition function." It is known that *all*

MRF's defined over neighborhood systems are Gibbs distributions and can, therefore, be expressed in terms of potential functions.

Denote by $\mathbf{d}_c$ a coefficient vector for the clique. (For example, the exponential argument in Equation 15 could be written using $\mathbf{R_f}^{-1} = \mathbf{d}_c\mathbf{d}_c^\dagger$.) We can apply a more general MRF model that allows us to rectify some of the limitations of the Gaussian prior by using the potential

$$V_c(\mathbf{f}) = \rho(\mathbf{d}_c^\dagger \mathbf{f}) \; . \tag{18}$$

The function ρ is monotone increasing but not necessarily convex. By relaxing the squared difference penalty we can reduce the too-high penalty applied to discontinuities. An example ρ sets $\rho(x) = \min\{|x|, T\}^2$ where T is a user-defined threshold. This choice has both computational and theoretical disadvantages: the nonconvexity of the model makes it difficult to globally minimize and the cost function is constant after the threshold is surpassed. This latter feature means that edges whose magnitude is larger than the threshold will be sharply reconstituted in the reconstruction while edges smaller than T will be smoothed.

The Huber function,

$$\rho(x) = \begin{cases} x^2 & \text{if} |x| \leq T \\ T^2 + 2T|x - T| & \text{if} |x| > T \, , \end{cases} \tag{19}$$

is convex and so avoids the problems associated with nonconvex ρ. Since the penalty for $|x| > T$ is *linear*, the excessive smoothing problems associated with quadratic penalties will also be reduced.

An (arguably) more useful MRF model applies variable weights w_c^2 to the quadratic penalty:

$$\rho(x) = w_c^2 \, x^2 \, . \tag{20}$$

Tikhonov's original suggestion,

$$\rho(\mathbf{d}_c^\dagger \mathbf{f}) = w^2(\mathbf{y}) \left[\left(\frac{\partial^2 f}{\partial y_1^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial y_1 \partial y_2} \right)^2 + \left(\frac{\partial^2 f}{\partial y_2^2} \right)^2 \right] , \tag{21}$$

is an example of this and enforces a strict smoothing constraint as the prior information. At a true image discontinuity, the weight $w^2(\mathbf{y})$ can be set to zero so as not to preclude the feature from the ensemble of allowed images. Moreover, the weights can be set to reflect the reliability to which this prior information is believed to hold.

This variable weight method generally results in a highly nonlinear (and possibly non-convex) reconstruction algorithm that often complicates numerical implementation. In addition, there are obvious problems with setting the magnitudes of the (location dependent) weights: the locations of the image discontinuities are usually part of the sought-for image information. The variable weight method is appealing, however, because it can be considered to be a more general version of Equation 18 than functions of the form of Equation 19. Moreover, the variable weight concept can be readily extended to "turn on/off" types of "constraint" criteria other than simple smoothing.

Let $\mathbf{w}$ denote the weight vector of Equation 20 so that $\mathbf{w}_i = w_c^2(\mathbf{y}_i)$. Choose

$$\int_{c\in C} V_c(\mathbf{f})\, dc = \mathbf{f}^\dagger \mathbf{d}_c \mathbf{W} \mathbf{d}_c^\dagger \mathbf{f}, \tag{22}$$

where $\mathbf{W}$ is the weight matrix whose main diagonal is determined by $\mathbf{w}$ (i.e., $\mathbf{W} = \mathrm{diag}\{\mathbf{w}\}$).

Substituting Equations 5 and 9 into Equation 14, with $p_{\mathbf{f}}(\mathbf{f};\mathbf{f}_0)$ defined by Equations 17 and 22, we have a representation of $B(p_{\mathbf{f}})$ in terms of the parameters $\mathbf{w}$:

$$\begin{aligned} B(\mathbf{W}) =& \|\mathbf{F} - \mathbf{K}\hat{\mathbf{f}}_{\mathrm{MV}}\|^2 + \mathrm{tr}[\mathbf{K}^\dagger \mathbf{K}(\mathbf{X}+\mathbf{R})^{-1}] \\ &+ \frac{\lambda}{2}\left(\hat{\mathbf{f}}_{\mathrm{MV}}^\dagger \mathbf{X}\hat{\mathbf{f}}_{\mathrm{MV}} - \ln\det[\mathbf{X}(\mathbf{X}+\mathbf{R})^{-1}] - \mathrm{tr}[\mathbf{R}(\mathbf{X}+\mathbf{R})^{-1}]\right) + \text{constant}, \end{aligned} \tag{23}$$

where $\mathbf{X} \equiv 2\mathbf{d}_c\mathbf{W}\mathbf{d}_c^\dagger$, $\mathbf{R} \equiv \mathbf{K}^\dagger\mathbf{R}_n^{-1}\mathbf{K}$, and $\hat{\mathbf{f}}_{\mathrm{MV}} = (\mathbf{X}+\mathbf{R})^{-1}\mathbf{K}^\dagger\mathbf{R}_n^{-1}\mathbf{F}$.

We are interested in minimizing $B(\mathbf{W})$ over the positive parameters $\mathbf{w}$. In general, these parameters may be chosen to be any value greater than zero and a general minimization algorithm can be developed in terms of gradients of $B(\mathbf{W})$ when $\mathbf{w}$ is allowed to vary in this "soft" way. Alternately, a "hard" choice could be made so that

$$\mathbf{w}_i = \begin{cases} w_{\mathrm{little}}, & \text{if the feature is to be left alone;} \\ w_{\mathrm{big}}, & \text{if the feature is to be readjusted.} \end{cases} \tag{24}$$

Such hard choices were employed by Reference 12 and effectively result in discontinuity detection. Various methods employing soft choices have also been proposed and yield somewhat more stable and robust techniques (c.f., References 14 and 15).

For illustration purposes in the present discussion, we shall employ the hard weights of Equation 24. The algorithm we use for choosing $\mathbf{w}$ is then very simple:

step 0: set $w_j = w_{\mathrm{big}}, \forall j$ and calculate $B_0 = B(\mathbf{W})$; set $i = 1$.

step 1: set $w_i = w_{\mathrm{little}}$ and calculate $\overline{B} = B(\mathbf{W})$.

step 2: if $\overline{B} < B_0$ then go to step 4.

step 3: set $w_i = w_{big}$

step 4: set $i \rightarrow i + 1$; if $i > M_1 M_2 = \dim \mathbf{W}$ then stop, otherwise go to step 1.

Once $\mathbf{X} = 2\mathbf{d}_c \mathbf{W} \mathbf{d}_c^\dagger$ is determined, then the estimate $\hat{\mathbf{f}}_{MV}$ is set.

Figure 1 demonstrates the method in a one-dimensional setting. The data consist of 25 uniformly distributed points that sample a simple function consisting of exactly two "steps." These data were contaminated with additive Gaussian white noise ($\sigma = 5\%$ of $\max(\mathbf{F})$) and are plotted in the figure. Also plotted in the figure are two 100-point (and so, "expanded") reconstructions developed from the data: the first estimate is the varying weight algorithm discussed above; and the second is the "ordinary" smoothed reconstruction with constant weights. The clique coefficient vector was chosen to be a second-order measure of smoothness with $\mathbf{d}_c$ the coefficients appropriate to a finite difference approximation to the second-order derivative: $\mathbf{d}_i^\dagger \mathbf{f} = f(y_{i+1}) - 2f(y_i) + f(y_{i-1})$. In the figure, $w_{big} = 100$, $w_{little} = 10^{-18}$, and $\lambda = 10^{-3}$.

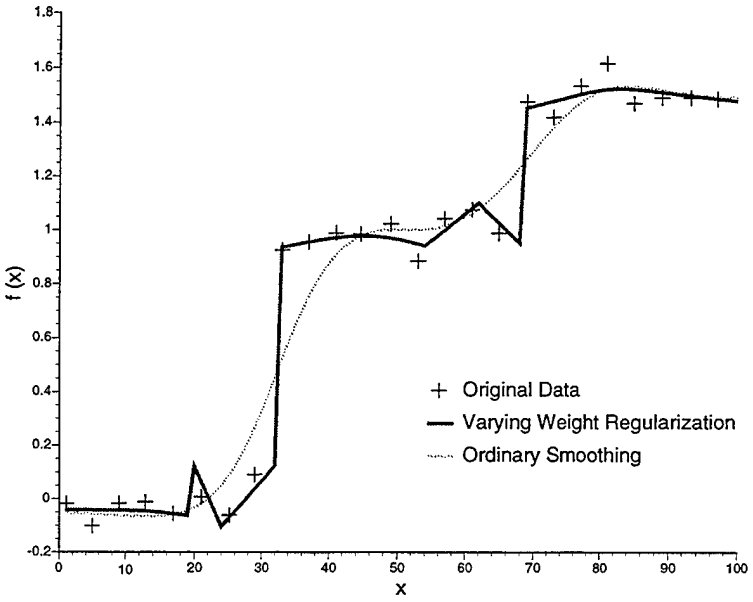

FIGURE 1. 4× Expansion Reconstruction of Simple Two-Step Function With $\sigma = 5\%$ of $\max(\mathbf{F})$, $\lambda = 10^{-3}$, $w_{big} = 100$, and $w_{little} = 10^{-18}$.

The example of Figure 1 is important in that it clearly demonstrates the difference between the current method and ordinary polynomial "fitting" techniques. The abrupt "jumps" in the reconstructed image—which correctly correspond to the associated jumps in the data—are not modeled by polynomial interpolates. Moreover, while the noise mitigation that occurs between the jumps is consistent with traditional methods, the ability to identify the jumps in the presence of noise is the focus of the current discussion.

ISAR IMAGING

ISAR is a method for increasing the resolution of a radar system by (effectively) increasing the size of the systems's aperture. When a radar target rotates in a known manner, then data collected as a function of time will also be data collected across a *synthetic* aperture. Careful assignment of these data to positions on this synthetic array often allow final image resolution that is many times that implied by the dimensions of the simple radar antenna.

In a monostatic scattering situation (in which the transmitter and receiver are co-located at range R and time t) it can be shown (Reference 18) that the weak scatterer approximation far-field response $H(\mathbf{k})$ due to a harmonic excitation of a target can be written:

$$H(k,\theta;R,t) \approx \frac{ikH_o\, e^{i(2kR-\omega t)}}{(2\pi)^2 R} \int_{\mathbb{R}^2} e^{i2ky'}\, e^{-i2kx'\theta}\, h(x',y')\, dx'\, dy'. \tag{25}$$

This is a linear superposition of waves radiating from a location on a plane whose normal is the (instantaneous) axis of rotation. The angle θ denotes the target aspect in the plane and we have employed the small angle approximation. Here, $h(\mathbf{r})$ is the local scatterer density function (scaled by range), ω is the angular frequency of the incident wave with: strength H_o; wave vector $\mathbf{k}$ and speed c. The parameters ω and c are related to $\mathbf{k}$ by $|\mathbf{k}| = \omega/c$, and the factor of 2 accounts for the two-way travel distance from the radar to the target and back again.

Equation 25 is a two-dimensional Fourier transform that can be readily implemented in digital systems. (Note, however, that the effect of approximating the polar data grid by the rectangular grid implied in Equation 25 is the introduction of image artifacts when x or y are large. We will not consider these artifacts further.) The ISAR Identity 25 neatly splits the target into down-range and cross-range factors. The quality of the target reconstruction will be determined by the size of the target resolution cell whose down-range and cross-range dimensions are inversely proportional to the bandwidth and angular aperture, respectively.

There are several practical difficulties in applying Equation 25 to the problem of estimating h, and these are discussed in the literature (References 18 and 19). The development of the Fourier transform relationship between $h(x,y)$ and $H(k,\theta)$ has relied on the approximation that h is independent of k and θ. This, of course, can only be completely true if the target consists of non-interacting and non-shadowing point scatterers—but the approximation is often justified by arguing that the bandwidth is sufficiently small and the angular aperture is sufficiently limited, so that h is "effectively" independent of k and θ for the duration of the measurement. And, in practice, this approximation is often sufficient for a good portion of the scattering elements (so-called "scattering centers") in a typical ISAR image. The scattering centers that do not accept this approximation consist of structural objects like ducts and inlets, corner reflectors, edges, and smooth surfaces whose specular return varies with aspect. In the presence of these complex scattering centers, an ISAR image generated by using Relation 25 will display "artifacts" characterized by streaks and curves, which often extend beyond the physical support of the target. Current research efforts carefully analyze these artifacts so that the type and actual location of the responsible scattering center might be estimated.

It is apparent that the central problem in modern ISAR imaging is one of obtaining enough resolution to separate the simple point scatterers from their more complex neighbors. In addition,

increased resolution is required for accurate imaging of the streaks and curves associated with the complex scatterers which, in turn, is necessary for follow-on image interpretation. In this sense, the ISAR imaging problem is quite similar to the image expansion problem already discussed: the effects of noise must be moderated; image point structure must be sharpened; and edges should not be blurred in the process. The central difference between the ISAR problem and optical image analysis is the complex values of the ISAR image and the (inherent) complex processing that must be performed for effective reconstruction.

For illustration purposes we will employ a two-dimensional second-order "measure of smoothness" that, in terms of the image coordinates, becomes

$$\int_{c\in C} V_c(\mathbf{f})\, dc = \mathbf{f}^\dagger \mathbf{d}_c \mathbf{W} \mathbf{d}_c^\dagger \mathbf{f} = \sum_{\substack{r=1,\ldots,M_1 \\ s=1,\ldots,M_2 \\ j=1,\ldots,4}} w^2(\mathbf{y}_{r,s})(\mathbf{d}_{r,s,j}^\dagger \mathbf{f})^2\,, \tag{26}$$

with the coefficient vector defined by

$$\begin{aligned}
\mathbf{d}_{r,s,1}^\dagger \mathbf{f} &= a_1 f(\mathbf{y}_{r,s+1}) + a_2 f(\mathbf{y}_{r,s}) + a_3 f(\mathbf{y}_{r,s-1}) \\
\mathbf{d}_{r,s,2}^\dagger \mathbf{f} &= b_1 f(\mathbf{y}_{r-1,s+1}) + b_2 f(\mathbf{y}_{r,s}) + b_3 f(\mathbf{y}_{r+1,s-1}) \\
\mathbf{d}_{r,s,3}^\dagger \mathbf{f} &= a_1 f(\mathbf{y}_{r-1,s}) + a_2 f(\mathbf{y}_{r,s}) + a_3 f(\mathbf{y}_{r+1,s}) \\
\mathbf{d}_{r,s,4}^\dagger \mathbf{f} &= b_1 f(\mathbf{y}_{r-1,s-1}) + b_2 f(\mathbf{y}_{r,s}) + b_3 f(\mathbf{y}_{r+1,s+1}),
\end{aligned} \tag{27}$$

(where $\mathbf{y}_{r,s}$ are image array coordinates). The choices $a_1 = a_3 = 1$, $a_2 = -2$, $b_1 = b_3 = 1/\sqrt{2}$, and $b_2 = -2/\sqrt{2}$ yield the finite-difference approximation to the second derivative of f (as in Equation 21). Here, however, we are applying a slightly more general form to allow for complex values of these coefficients.

Figure 2 is an example of this method applied to synthetic ISAR data. The original $M_1 \times M_2 = 24 \times 24$ scene is displayed in 2a. An $N_1 \times N_2 = 16 \times 16$ data array was generated by under-sampling the array $\mathbf{Kf}$ generated from the original scene. This, of course, results in reconstruction artifacts, which are displayed in 2b. (Figure 2b was generated by $(\mathbf{K}^\dagger\mathbf{K})^{-1}\mathbf{K}^\dagger\mathbf{F}$ with $\mathbf{F}$ the noise-free undersampled data.) Then a noisy undersampled data set was constructed using $\sigma = \frac{1}{2}\sqrt{\|\mathbf{F}\|^2/N_1N_2}$. The corresponding "traditional" reconstructed image (as in 2b) is shown in Figure 2c. Figure 2d is the variable weight reconstruction with $\lambda = 100$, $w_{\text{big}} = 10$, and $w_{\text{little}} = 10^{-3}$.

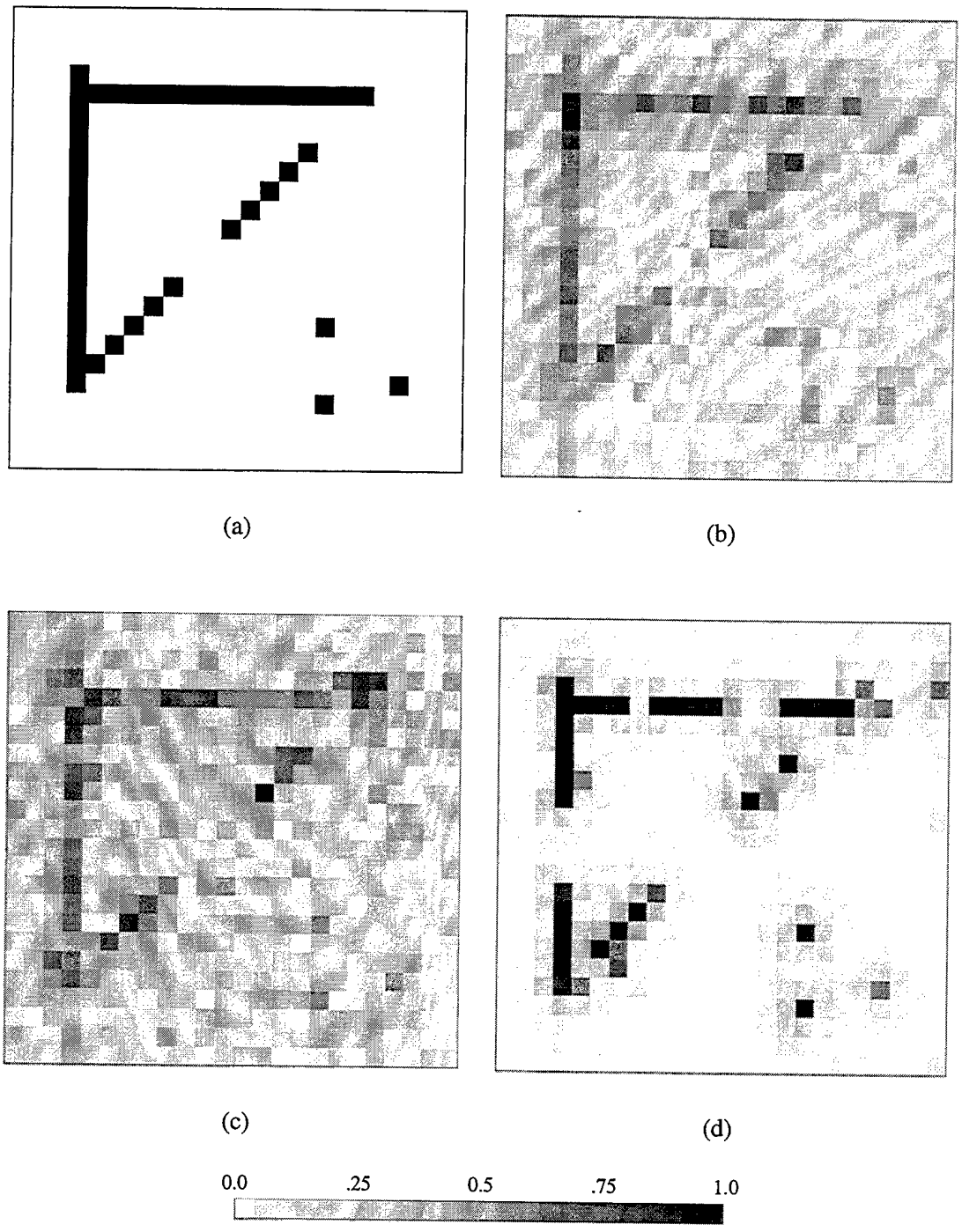

FIGURE 2. Example ISAR Image Reconstruction. (a) Original $M_1 \times M_2 = 24 \times 24$ scene; (b) Simple (Fourier) image reconstructed from an undersampled ($N_1 \times N_2 = 16 \times 16$) data set $\mathbf{F}$ *without noise* (to explicitly show undersampling artifacts); (c) Image reconstructed from noisy undersampled data (with $\sigma = \frac{1}{2}\sqrt{\|\mathbf{F}\|^2/N_1N_2}$); and (d) Variable weight reconstruction with $\lambda = 100$, $w_{\text{big}} = 10$, and $w_{\text{little}} = 10^{-3}$.

The smoothing constraint that we applied to the data set in the example of Figure 2 was *ad hoc* and constructed in an effort to reduce the effect of both sidelobes and image noise. In Equation 27 we set

$$\begin{aligned} a_1 &= e^{i\pi\left(\frac{1}{2}-\frac{N_1-1}{M_1}\right)}, \qquad a_2 = -2, \qquad a_3 = e^{i\pi\left(\frac{1}{2}+\frac{N_1-1}{M_1}\right)} \\ b_1 &= \frac{1}{\sqrt{2}}e^{i\pi\left(\frac{1}{2}-\frac{N_2-1}{M_2}\right)}, \qquad b_2 = -\frac{2}{\sqrt{2}}, \qquad b_3 = \frac{1}{\sqrt{2}}e^{i\pi\left(\frac{1}{2}+\frac{N_2-1}{M_2}\right)} . \end{aligned} \tag{28}$$

This has the effect of rotating the sidelobes associated with an image element by $\pi/2$ with respect to the local peak of the element—and at the same time does not "over emphasize" the variations due to noise.

DISCUSSION AND CONCLUSION

Bayesian (and associated) methods that assign probability density to image "intensity" can be justified in many circumstances. When the data are complex-values, however, or when it is allowed to take on negative values, these kind of probability assignments may not be useful because they discard important aspects of the problem (e.g., phase). In the foregoing discussion we have addressed this issue by associating the relevant probabilities to an image ensemble that is defined by the noise process and the *a priori* information available. This method is generally independent of data "type" considerations and, in particular, is not associated with image intensity.

By freeing the method from an artificial dependence on data amplitude we are able to build a more general method that allows us to compare the local phase values of a reconstructed (complex-valued) image. Our approach has been for a regularizational functional consisting of two terms: the first of which models the effects of noise and measures the "energy" difference between the data and the reconstructed image; and a second term that tracks the "information" difference between the reconstructed image and a prior image model. This model can be used to enforce smoothness constraints and we have given several (elementary) examples to illustrate how that can be accomplished. These examples are simple and easy to implement, but the smoothness criteria may not be optimal for ISAR image reconstruction. Moreover, our implementation of the functional "minimization" required to determine the minimizing (estimate) image was a simple expedient that involved a binary variation across the image elements and can almost certainly be improved upon.

Our current efforts are concentrating on: determining which (phase dependent) functionals are appropriate to more general ISAR imaging and extending the optimization search algorithm to allow for continuous variation in $\mathbf{W}$.

REFERENCES

1. K. M. Hanson. "Bayesian and Related Methods in Image Reconstruction From Incomplete Data," in *Image Recovery: Theory and Practice,* ed. by Henry Stark. Orlando, Fla., Academic Press, 1987.

2. A. N. Tikhonov and V. Y. Arsenin. *Solution of Ill-Posed Problems.* New York, Wiley, 1977.

3. D. Keren and M. Werman. "Variations on Regularization," in *Proceedings of the 10th International Conference on Pattern Recognition, 16–21 June 1990, Atlantic City, N.J.* Washington D.C., IEEE Computer Society Press, 1990, Pp. 93–8.

4. D. M. Titterington. "The Maximum Entropy Method for Data Analysis," *Nature,* Vol. 312, No. 22 (1984), pp. 381–2.

5. V. A. Macaulay and B. Buck. "Linear Inversion by the Method of Maximum Entropy," *Inverse Problems,* Vol. 5, No. 5 (1989), pp. 859–74.

6. B. R. Frieden and A. T. Bajkova. "Bayesian Cross-Entropy of Complex Images," *Applied Optics,* Vol. 33, No. 2 (1994), pp. 219–26.

7. A. P. Sage and J. L. Melsa. *Estimation Theorey with Applications to Communications and Control.* Huntington, Del., Robert E. Krieger, 1979.

8. I. Csiszár. "Why Least Squares and Maximum Entropy? An Axiomatic Approach to Inference for Linear Inverse Problems," *Annals of Stat.,* Vol. 19, No. 4 (1991), pp. 2032–66.

9. J. E. Shore and R. W. Johnson. "Axiomatic Derivation of Maximum Entropy and the Principle of Minimum Cross-Entropy," *IEEE Trans. Inf. Theory,* Vol. IT-26, No. 1 (1980), pp. 26–37.

10. T.J. Cornwell. "Is Jaynes' Maximum Entropy Principle Applicable to Image Construction?" in *Indirect Imaging: Measurement and Processing for Indirect Imaging,* ed. by J. A. Roberts. Cambridge, Mass., Cambridge University Press, 1984.

11. A. Mohammad-Djafari and G. Demoment. "Estimating Priors in Maximum Entropy Image Processing," in *International Conference on Accoustics, Speech, and Signal Processing, 3–6 April 1990, Albuquerque, N.Mex.* Piscataway, N.J., IEEE Service Center, 1990, Pp. 2069–72.

12. D. Lee and T. Pavladis. "One-Dimensional Regularization with Discontinuities," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* Vol. 10, No. 6 (1988), pp. 822–9.

13. D. Terzopoulos. "Regularization of Inverse Visual Problems Involving Discontinuities," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* Vol. 8, No. 4 (1986), pp. 413–24.

14. S. S. Sinha and B. G. Schunck. "Discontinuity Preserving Surface Reconstruction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4–8 June 1989, San Diego, Calif.* Piscataway, N.J., IEEE Service Center, 1989, Pp. 229–34.

15. H. C. Andrews and B. R. Hunt. *Digital Image Restoration.* Englewood Cliffs, N.J., Prentice-Hall, 1977.

16. R. L. Stevenson, B. E. Schmitz, and E. J. Delp. "Discontinuity Preserving Regularization of Inverse Visual Problems," *IEEE Trans. Systems, Man, and Cybernetics,* Vol. 24, No. 3 (1994), pp. 455–69.

17. R. R. Schultz and R. L. Stevenson. "A Bayesian Approach to Image Expansion for Improved Definition," *IEEE Trans. Image Proc.,* Vol. 3, No. 3 (1994), pp. 233–42.

18. D.L. Mensa. *High-Resolution Radar Imaging.* Dedham, Mass., Artech House, 1981.

19. A. W. Rihaczek and S. J. Hershkowitz. *Radar Resolution and Complex-Image Analysis.* Dedham, Mass., Artech House, 1996.

INITIAL DISTRIBUTION

4 Naval Air Systems Command, Arlington
- AIR-4.1.1 (1)
- AIR-4.1.8
  - File (1)
  - Stegman (1)
- AIR-4.10A (1)

2 Defense Technical Information Center, Fort Belvoir
2 Center for Naval Analyses, Alexandria, VA
- Dr. A. Borden (1)
- L. Lynn (1)

ON-SITE DISTRIBUTION

1 Code 455560D, W. Katzenstein
1 Code 455580D, D. Reade
2 Code 4713C0D
- G. Hewer (1)
- R. Smith (1)

1 Code 472230D, D. Paolino
1 Code 472G70D, G. Winkler
1 Code 473A20D, M. Mumford
2 Code 474400D
- B. Borden (1)
- S. Chesnut (1)

1 Code 474T60D, T. Loftus
4Code 4BL000D (3 plus Archives copy)